Anomaly Density Estimation from Strip Transect Data: Pueblo of Isleta Example

Sean A. McKenna, Sandia National Laboratories
Brent Pulsipher, Pacific Northwest National Laboratory

May 2005

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAY 2005		**00-00-2005 to 00-00-2005**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Anomaly Density Estimation from Strip Transect Data: Pueblo of Isleta Example	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Pacific Northwest National Laboratory,902 Battelle Boulevard,Richland,WA,99354**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**19**	

Introduction

Extensive geophysical surveying done at a wide variety of sites across the country has shown that on any one site, the density, or intensity, of geophysical anomalies varies from one location to another. This spatial variation makes it difficult to determine the total number of anomalies at any one location given that it may not be possible to geophysically survey the entire site. However, it is desirable to accurately estimate the number of anomalies at specific locations in order to determine boundaries of target locations or other regions where the anomaly density is above the background density. Also, an accurate picture of the anomaly distribution is required for pre excavation cost estimation and planning worker safety measures.

This report documents an approach to estimating spatially variable anomaly density from limited transect data. This approach is then tested using a data set collected at the Pueblo of Isleta S1 site in central New Mexico. The data set was collected by the Naval Research Laboratory using an airborne magnetometer array (NRL, 2005). The data set provides an exhaustive picture of the anomaly locations over an area of nearly 1500 acres at the Isleta site. Excavation at this site has identified UXO in several locations and the entire 1500 acres may be considered as a target area. This study on anomaly density estimation does not attempt to classify the site into different types of areas (e.g., background vs. target area) nor establish boundaries for potential target areas. The work here is solely focused on estimating the number of anomalies at all locations. Three different transect sampling designs covering 2.85, 0.90 and 0.45 percent of the total site are used to provide the data for the estimation. The results of the density estimation are evaluated by comparing the estimates to the true surveyed anomalies at the unsampled locations.

The anomaly density estimation approach documented herein utilizes a spatial Poisson model for the distribution of the anomalies. Two sources of uncertainty in the application of this Poisson model are identified but are not fully exercised at this stage. The means of extending the approach developed here to incorporate these uncertainties and provide a full distribution of the uncertainty in the final estimates of the anomaly density is outlined in this report, but application of this extended approach is left for future analyses.

Estimation Approach

The steps of the anomaly density estimation approach and the inputs necessary for each step are outlined in Table 1. There are five main steps necessary to obtain the data and produce the final estimates of the anomaly density. These steps are discussed below.

Table 1. Steps in the anomaly estimation procedure

Step	Necessary Inputs
1) Transect Design	Site Specific DQO's, CSM
2) Definition of Modeling Grid	DQO's, geophysical transect width, CSM
3) Upscaling of transect data to model grid cells	Transect data, DQO's
4) Variogram estimation	Transect data upscaled to grid cells
5) Kriging of cell intensity values across site	Transect data upscaled to grid cells, variogram model

Step 1

The source of data for either target detection or density estimation is a set of geophysical transects. The design of these transects is based on a set of data quality objectives (DQO's) and a conceptual site model (CSM). The DQO's and CSM are used with the algorithms in the Visual Sampling Plan (VSP) software to determine the required transect design. Typically, the transects are linear, parallel transects with a fixed transect width and a fixed spacing between the transects. These geophysical transects provide the X,Y locations of the anomalies that fall within them. Typically, the anomaly signal strength is also provided as part of the transect data, although this signal strength is not currently used in the anomaly density estimation.

Step 2

By definition, the anomaly *density* is the number of anomalies for a given area. Therefore it is necessary to define the area over which the number of anomalies will be estimated. This area could simply be the entire site, but more commonly, the entire site is discretized into a number of equal area cells and an estimate of the number of anomalies within each cell is required. This discretization makes it possible to make excavation decisions and cost estimates for each cell providing a much more finely resolved picture of the site conditions as compared to the case of just a single estimate for the entire site.

The size of the grid cells needs to consider the DQO's, the geophysical transect width and the CSM. In the next step, the transect data will be upscaled from the scale of the transects to the scale of the model grid cells. The goal in this upscaling is to keep the area over which the transect data are averaged consistent with the area of the grid cells. Additionally, the grid cells have to be of a size that will allow specific aspects of the CSM to be identified in the final estimated values –cells that are too large cells will not provide detail on features identified in the CSM while cells that are too small will result in high cell to cell variance in the final estimates as they do not average over enough of the site area. The relationship between the grid cell size and the transect size is discussed in the next step.

Step 3

Previous work has shown that simply laying the grid cells on top of the transect data and assigning the transect values to the corresponding grid cell does not provide consistent

estimates of the anomaly density from one cell to the next. This problem is exacerbated when relatively narrow transects are used and/or the background anomaly density of the site is low (i.e., only a few anomalies per acre). A solution to this problem is to calculate a spatial average of the anomaly density using transect data that lie outside of the current modeling grid.

The spatial averaging approach used in this work is shown schematically in Figure 1. The average intensity assigned to the red cell in the center of Figure 1 is calculated as the average of the intensity values along the transect within an averaging window, shown as the red segments of the transect in Figure 1. The cell to which the average intensity is assigned lies at the center of the averaging window. The width of the transect is typically much smaller than the dimensions of the grid cells. However, by averaging over several segments of the transect, the total area in the calculation of the average can approximate the area of the grid cell. For example, for grid cells that are 50x20m ($1000m^2$) and a transect that is 3m wide, each 50 m long segment of the transect will be define an area of $150m^{2.}$ If an averaging window containing 7 segments is used, the intensity will be calculated over an area of $7x150m^2$, or $950m^2$, which is very close to the cell area of $1000m^2$.

This averaging approach assumes that the average intensity calculated along a section of the transect is representative of the intensity within the grid cell at the center of the averaging window if it were possible to completely sample that cell. This assumption is reasonable in that the length of the averaging window and the grid cell size can be chosen to minimize the difference in area between them. However, the length of the averaging window must be set such that it is long enough to provide representative samples of the anomaly intensity, but not be too long such that areas of the transect far from the center cell can unduly influence the calculation of what needs to be a local average. The transect spacing and/or the expected target area size can be used to guide the choice of averaging window length.

The averaging process produces a data set defined at the center of the grid cells that were intersected by the transects. The primary information obtained in this data set is the average, or expected, number of anomalies that would be found in that grid cell. The expected number of anomalies divided by the cell area define the intensity of a Poisson process at that cell location. More specifically, these data define a non-homogeneous Poisson process (NHPP) as the intensity is not constant from one cell to the next. The next step is to estimate the intensity of this NHPP at all cells that were not sampled by the transects.

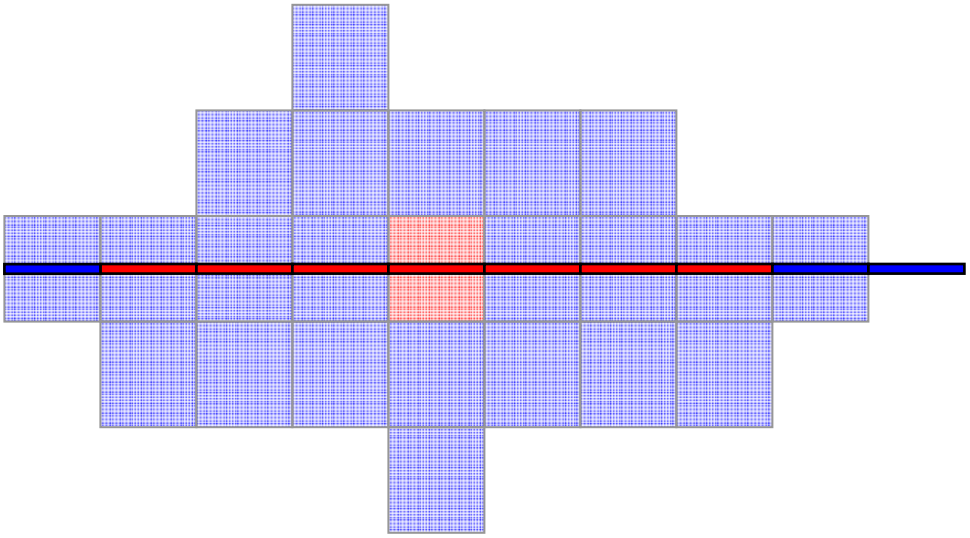

Figure 1. Schematic diagram of the averaging process to estimate grid cell intensity values from transect data

Step 4

The expected value of the anomalies in each cell that is intersected by a transect provide the data for the calculation of the variogram. The variogram defines the spatial variation in the anomaly intensity across the site as calculated from the transect data. Due to the averaging process to upscale the transect data to the grid-cell scale, the spatial variation of the anomaly intensity will be smoothed, or "regularized", to some degree. This regularization process results in variograms with nugget values close to zero and often a Gaussian variogram model provides the best fit to the points calculated as the experimental variogram. The important point to understand in this step is that the variogram produced here describes the spatial variation of the anomaly intensity as modeled by a NHPP measured at the scale of the grid cell. It is not a description of the spatial variation that would occur at the much smaller scale of the geophysical transect.

Step 5

The spatial estimation of the NHPP intensity is done using the geostatistical estimation algorithm of ordinary kriging (OK). The variogram model fit to the experimental variogram provides a measure of spatial covariance between any two points separated by a straight-line distance. The estimation proceeds by calculating the intensity at each unsampled location as a weighted linear average of surrounding data points. The weights are calculated to provide an unbiased, minimum variance estimate of the intensity at the unsampled location through solution of the OK system. These estimates are made using the covariances between the cells that have sample data as well as the covariances

between these cells and every cell where an intensity estimate is required (the "unsampled" cells). Estimates are made at all unsampled grid cells within the site domain.

At the completion of the spatial estimation step, every grid cell contains an estimate of the Poisson intensity. The Poisson distribution is noteworthy in that it is a single parameter distribution with both the mean and the variance equal to the intensity value. While the intensity values, and thus the mean and variance, of the Poisson distribution are continuous variables, the Poisson distribution itself is discrete. The Poisson distribution is given as:

$$P(X) = \frac{e^{-\lambda}\lambda^{X}}{X!}$$

Where λ is the intensity (objects/area), and X is the number of objects. $P(X)$ is the probability of exactly X objects existing in that area given the intensity, λ. To provide a feel for the Poisson distribution, histograms of three Poisson distributions with intensity values of 0.2, 2 and 20 are shown in Figure 2.

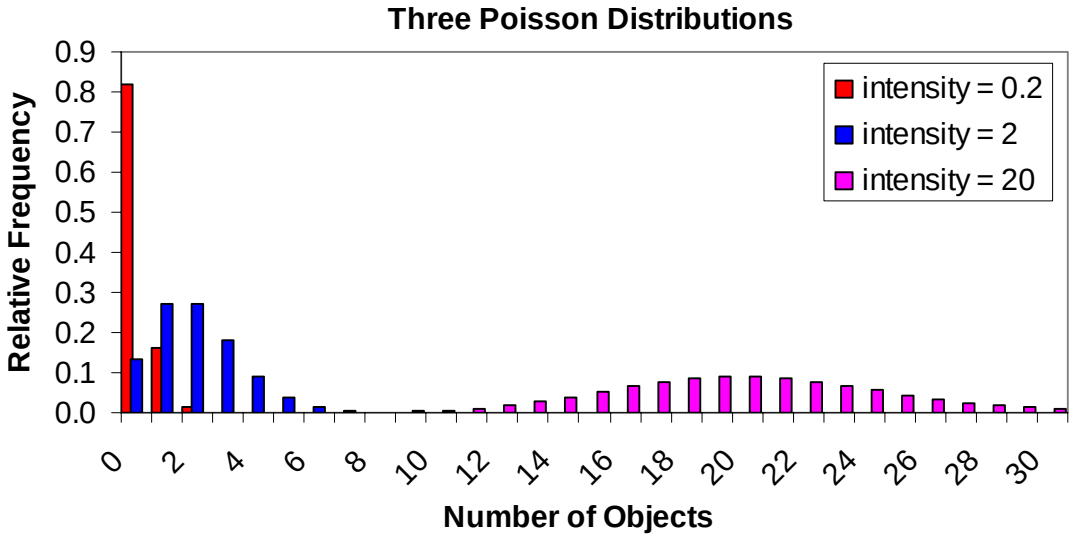

Figure 2. Example Poisson distributions with different intensity values.

As an example, the probability of having exactly 2 anomalies in a grid cell given that the intensity estimated for the grid cell is 0.2, 2 or 20 is 1.6E-02, 0.27 and 4.1E-07 respectively. These examples demonstrate that knowledge of the Poisson intensity at a location does not uniquely determine the number of anomalies at that location, but provides the distribution that defines the uncertainty in the number of anomalies at that location. For this work, we simply report the mean intensity of the Poisson distribution as estimated at each grid cell. An extension to this approach to account for the full Poisson distribution at each grid cell is discussed below.

Extensions to Quantify Uncertainty

The approach outlined above was designed to produce estimates of the anomaly intensity at each grid cell and to account for the uncertainty in these estimates. At the present time, this uncertainty is not sampled and only a single deterministic value for the number of anomalies within each cell is estimated. However, the approach to anomaly density estimation was developed so that it could be easily extended to incorporate two major sources of uncertainty. These two major sources of uncertainty are: 1) The uncertainty in the spatially varying Poisson intensity as estimated by the kriging algorithm for any given location; and 2) The actual number of anomalies within a cell for a specified intensity.

The uncertainty in the spatial distribution of the Poisson intensity can be quantified by *simulating* the spatial varying intensity. At the current time we are *estimating* the spatially varying intensity values. Estimation produces both an expected value of the intensity as well as a variance about that expected value. Simulation applies Monte Carlo sampling to the distribution defined by the estimation procedure. This Monte Carlo sampling is done in such a way that a series of stochastic intensity maps are produced. Each of these maps honors the observed intensity data at the grid cells that were intersected by a transect as well as honoring the histogram and variogram of those cell data. Multiple stochastic simulations can be produced so that at every grid cell a full distribution of possible intensity values will be defined. The average of all these simulated intensities will be equal to the single estimated value that is being produced in the current approach.

As shown in Figure 2, each estimated, or simulated, intensity fully defines the Poisson distribution at that location. Therefore, for a given intensity value, multiple Monte Carlo realizations can be drawn from the corresponding Poisson distributions to provide an uncertainty distribution on the total number of objects within an area. At this time, we are simply reporting the average of this distribution.

Monte Carlo techniques can be used to sample both the spatial distribution of intensity as well as the Poisson distribution defining the total number of anomalies at each location. Geostatistical simulation algorithms are available for the spatial simulation and software to do Monte Carlo sampling of the Poisson distribution has already been written for this project. While this level of detail in uncertainty quantification may be beyond the need of the field user, it would be worthwhile to complete some additional calculations and determine which type of uncertainty, spatial vs. Poisson, has the greatest effect on the final anomaly count estimates. Those calculations are saved for a later date and not done here.

Data Set

The Pueblo of Isleta S1 Site in central New Mexico is used to demonstrate the NHPP approach to estimating anomaly intensities as described above. Magnetometer data collected with an airborne system by the Naval Research Laboratory serve as the ground truth for this study (Figure 3). The entire approximately 1500 acre survey area is considered to be within the target region of the S1 site. A large area of dense anomaly intensity near the south central portion of the surveyed area is evident in Figure 3.

Additional small clusters as well as individual anomalies scattered throughout the site area. Several linear features of high anomaly density (e.g., the north-south feature at an X coordinate of approximately 9000 m in Figure 3) in the surveyed area are most likely fences and/or pipes. However, the exact cause of these linear features was not identified and the anomalies observed along them are not treated any differently than any other anomalies in the data set. The total number of anomalies in the surveyed area is 16,335. The coordinate system used in this work is modified from the original UTM coordinate system by subtracting 310,000 from the X coordinates and 3,850,000 from the Y coordinates.

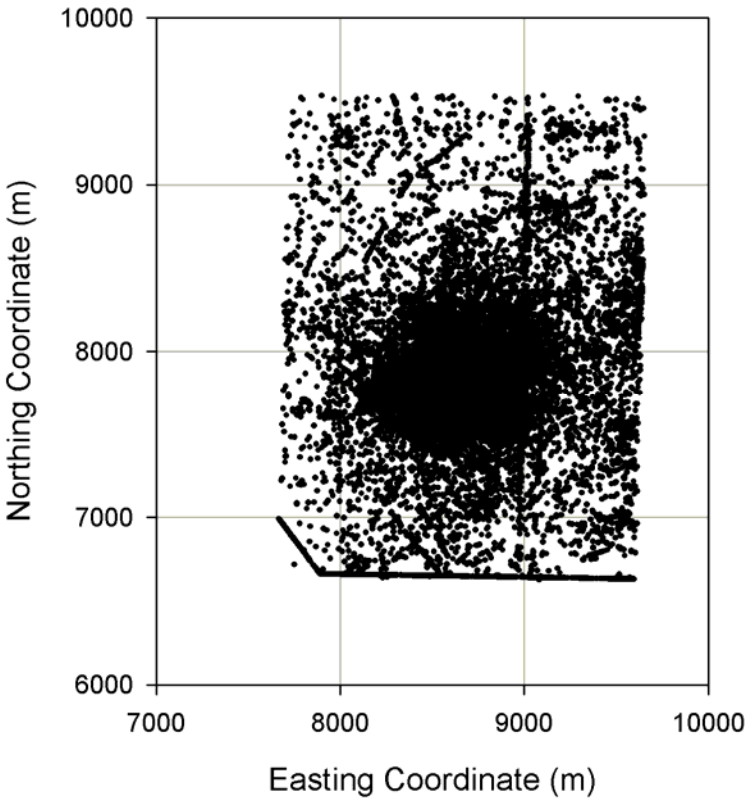

Figure 3. Location of geophysical anomalies as identified through airborne surveying.

The survey area is trimmed for this study to provide a rectangular study region with boundaries that are oriented north-south and east-west. The east-west extent of the trimmed boundary in the modified coordinates is from coordinates 7660 to 9660 (2000m) and the north-south extent is from coordinates 6675 to 9525 (2850 meters). The total number of surveyed anomalies in this trimmed area is 16,033. The site is sampled using three different transect sampling designs. The transect locations, as overlain on the anomaly data, are shown in Figure 4. The three transect spacings used are 105.42, 316 and 631.7 meters as shown in the top, middle and bottom images of Figure 4 respectively. Each transect is three meters wide and the three transect designs cover 2.85, 0.9 and 0.45 percent of the site, respectively.

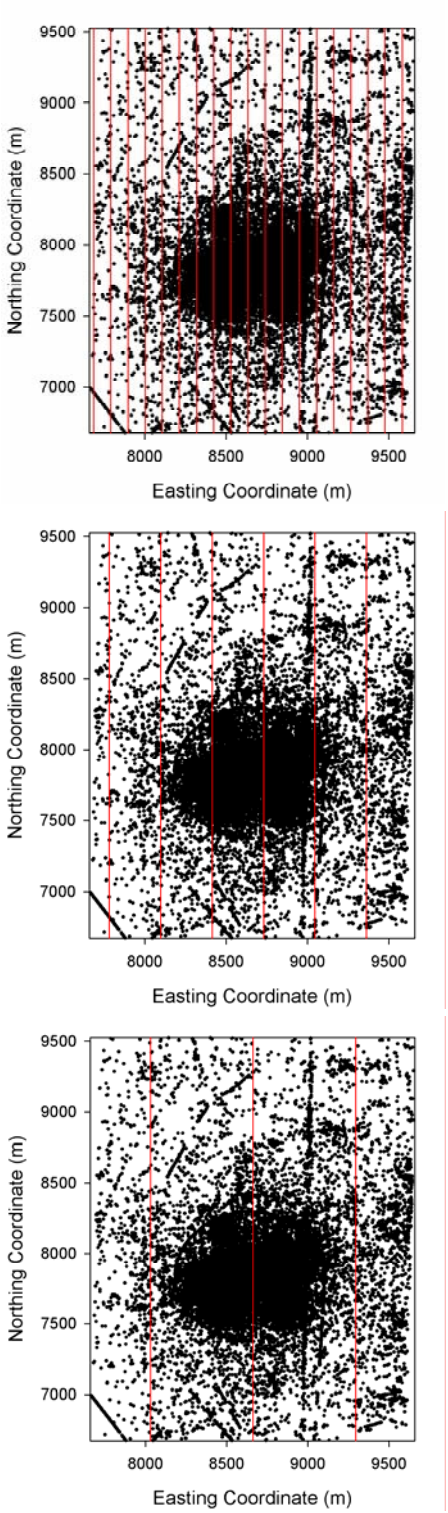

Figure 4. Three sampling patterns (105, 316 and 631 meter spacing from top to bottom, respectively) overlain on the anomalies identified through airborne geophysical surveys.

The locations of the transects, the number of anomalies identified on each transect and the anomaly intensity per acre are shown in Table 2 for each of the three transect spacings. The total number of anomalies identified on the transects is 475, 164 and 69 for the 105, 316 and 631 meter spacings respectively.

Table 2. Transect coordinates and sampling results

Transect Spacing	Transect ID	Transect X Coordinate	Number of Anomalies	Intensity (count/acre)
105.42 meters	1	7580.4	4	1.9
	2	7685.8	6	2.8
	3	7791.3	3	1.4
	4	7896.7	5	2.4
	5	8002.1	10	4.7
	6	8107.5	16	7.6
	7	8212.9	28	13.3
	8	8318.4	48	22.7
	9	8423.8	70	33.1
	10	8529.2	46	21.8
	11	8634.6	48	22.7
	12	8740.0	52	24.6
	13	8845.5	40	18.9
	14	8950.9	35	16.6
	15	9056.3	16	7.6
	16	9161.7	8	3.8
	17	9267.1	15	7.1
	18	9372.6	11	5.2
	19	9478.0	15	7.1
316 meters	1	7782.0	8	3.8
	2	8098.0	11	5.2
	3	8414.0	45	21.3
	4	8730.0	64	30.3
	5	9046.0	22	10.4
	6	9362.0	12	5.7
631.7 meters	1	8031.7	7	3.3
	2	8663.4	49	23.2
	3	9295.1	13	6.2

The study domain is covered with a modeling grid of 100 by 57 cells. Each cell is 20m wide (X direction) and 50m tall (Y direction) for an area of $1000m^2$ per cell. An expected Poisson intensity is determined for each cell that is intersected by a transect using the averaging procedure described above. The averaging window along the transect is 7 cells, or 350m, long. The average intensity values along the transects in anomalies per cell (anomalies per $1000m^2$) are shown in Figure 5.

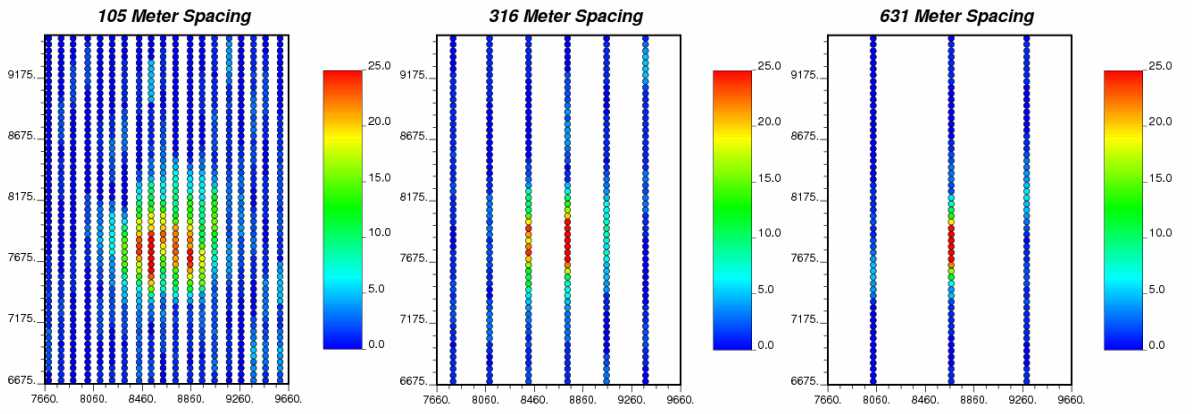

Figure 5. Expected intensity values (anomalies/1000m^2) along the three different sets of transects

Summary statistics of the intensity values as determined by the moving window averaging process are shown in Table 3. In general, these statistics show that the distribution of intensity values has a positive skew, as exhibited by the mean value being larger than the median, and that a large proportion of the expected intensity values are zero (note that the 25th percentile is zero for the 105 and 631m spacing transect data). There is quite a bit of variation in the sample statistics across the three different transect patterns. The statistics of the 105m spacing transects are the closest to the statistics of the true data set (right column of Table 3).

Table 3. Statistics of expected intensity values at grid cells intersected by transects. The same statistics for the true (exhaustive) data set are also shown. All units are in anomalies per 1000m^2 cell.

Statistic	105m Transects	316m Transects	631m Transects	True
Mean	2.89	3.13	2.65	2.81
Std. Deviation	5.10	5.86	5.75	5.40
Minimum	0.0	0.0	0.0	0.00
25th percentile	0.0	0.95	0.0	0.00
Median	0.95	0.95	0.95	1.00
75th percentile	2.86	2.86	1.90	3.00
Maximum	33.33	34.29	30.48	39.00
Num. Samples	1083	342	171	5700

The expected intensity values shown in Figure 5 are used as input to the variogram calculation and modeling process. The experimental variograms calculated and the models fit to those experimental variograms are shown in Figure 6. In all cases, a Gaussian model was fit to the experimental variograms. For each model, the sill of the variogram, maximum gamma value of the model, was fixed to the variance of the data set. This is the theoretically appropriate way to identify the sill value. The experimental points that lie above the sill exhibit a "hole effect" and are due to the local area of high intensity values in the south-central portion of the site.

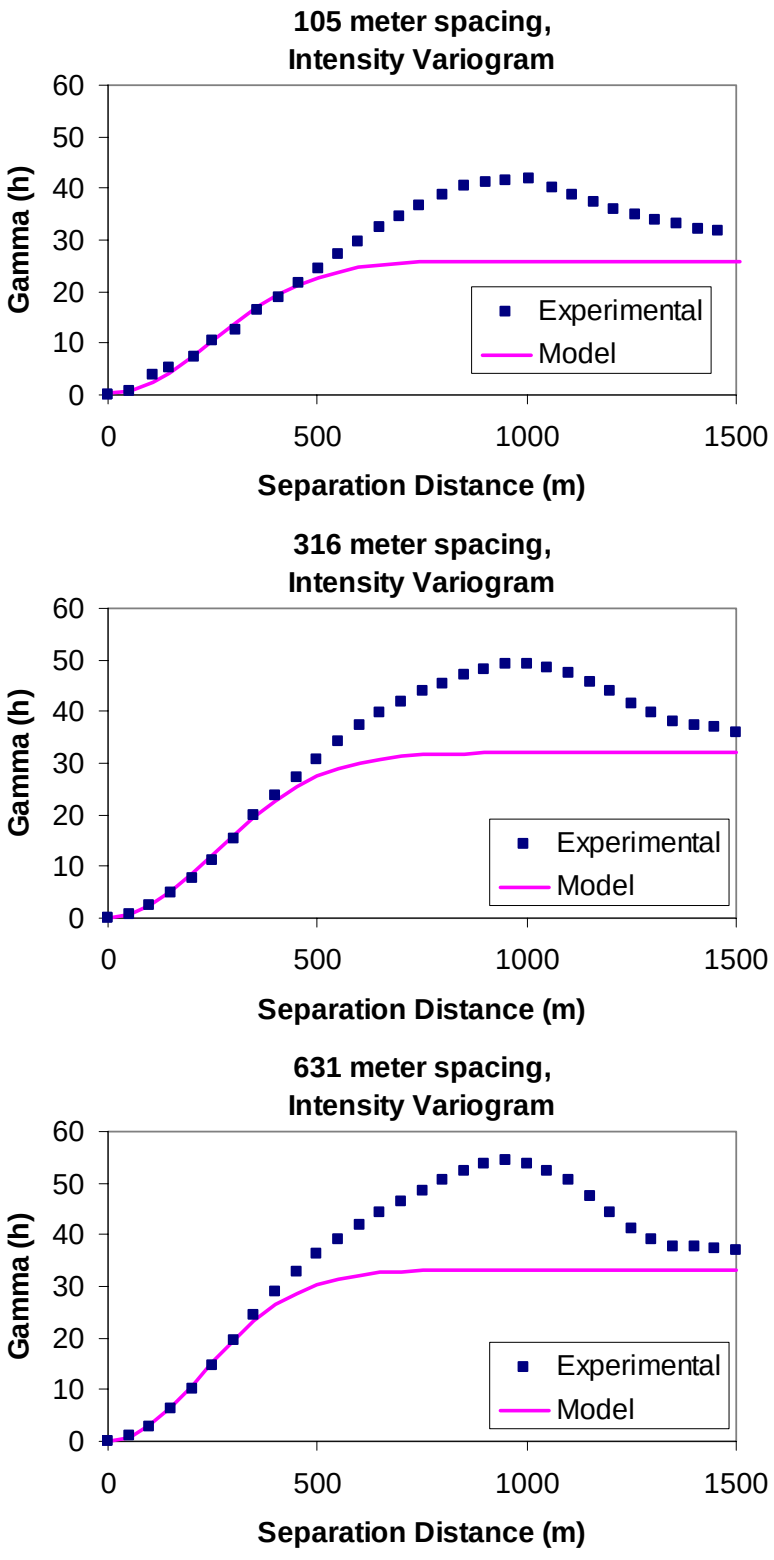

Figure 6. Experimental variograms and the models fit to them for the three different sampling designs.

The variogram models are similar across all transect designs with ranges of 1050, 1075 and 950 meters and sills of 26.0, 32.0 and 33.1 (anomalies per cell)^2 for the 105, 316 and 631 meter spacing transect spacing designs, respectively.

The next step in the density estimation is to estimate the Poisson intensity at all unsampled cells in the model grid. This is done using the three different data sets as input to the OK algorithm and the corresponding variograms shown in Figure 6. The result is an estimate of the number of anomalies within each cell in the model grid. These estimates are shown as maps in Figure 7. The upper left image of Figure 7 also shows the true number of anomalies per cell for comparison.

Visual inspection of the results in Figure 7 shows that the kriging estimates of the intensity values are capable of reproducing the large scale characteristics of the true intensity field, but are not capable of reproducing the fine-scale features in the true field. The kriging estimates represent smoothed versions of the true field and, as expected, the degree of smoothing increases as the transect spacing increases. Fine scale features in the true field that are parallel or oblique to the transect direction are difficult to estimate. Examples of these include the nearly north-south higher intensity feature that intersects the east side of the large high intensity region in the south central portion of the study area and the diagonal feature in the southwest corner of the study area. Anomaly intensity within large features without a preferred orientation, such as the main high density region in the south central portion of the study area, are better reproduced by the kriging.

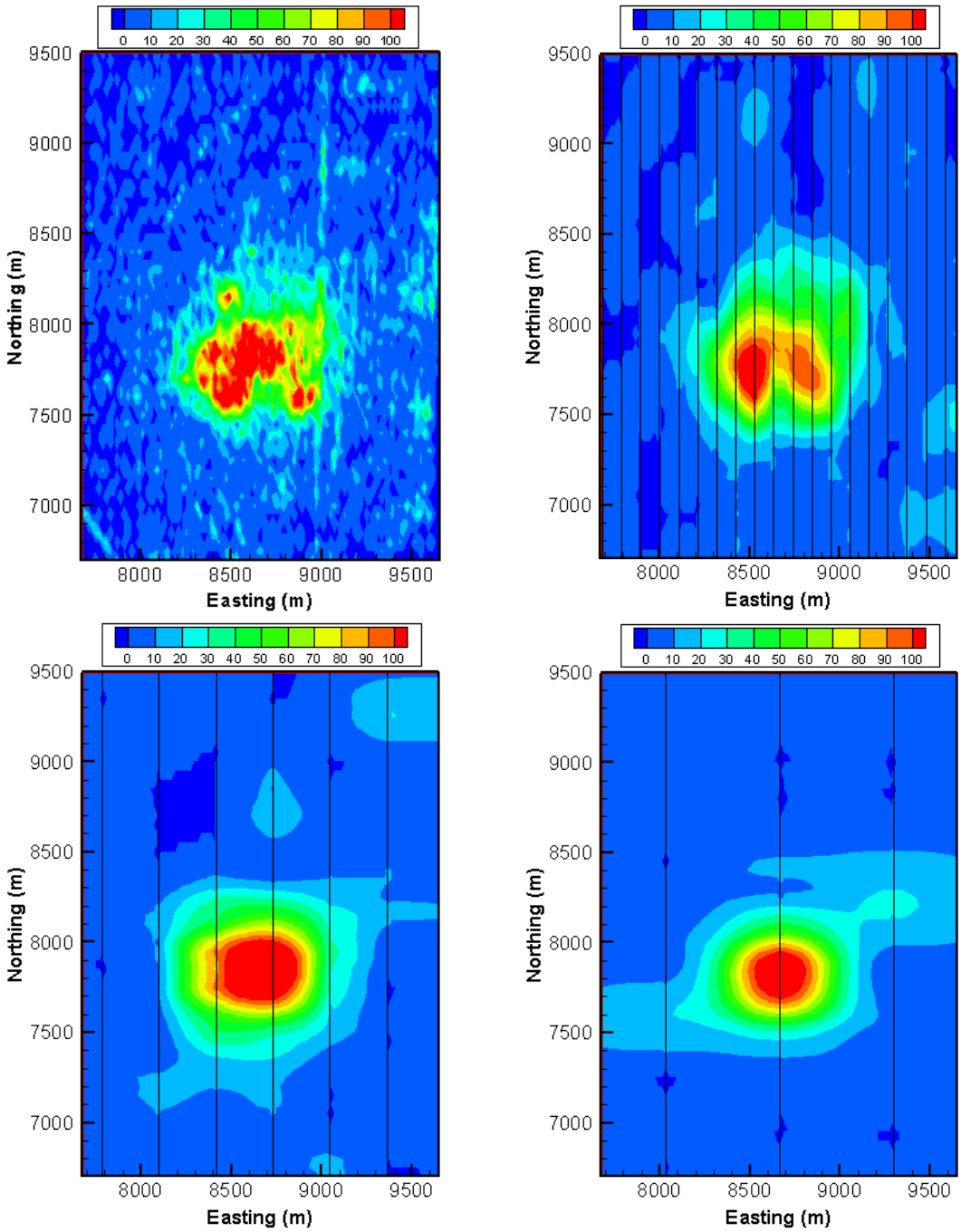

Figure 7. Estimated anomaly densities for the true field (upper left) and the 105, 316 and 631 m spacing transect designs. The color scale shows the true or estimated number of anomalies per acre within each 1000m^2 grid cell.

In addition to the visual inspection of the results, each cell in each of the three estimated fields is compared to the true cell values more quantitatively. For each cell, the error between the estimated and the true number of anomalies in the cell is calculated as: (*estimated-true)* such that overestimates produce positive errors and vice-versa. Ideally, the estimates will be unbiased producing a mean error of zero and have a relatively small variance about that mean. The errors are evaluated for two different sets of grid cells: 1)

those where the transect crossed the grid cell and the error is due solely to the upscaling process to go from the three-meter wide transect to the 20 meter wide grid cell; and 2) the cells that did not have a transect intersecting them. At these cells, the error is due to the geostatistical estimate of the intensity value. These errors also incorporate the upscaling errors as the sample values were used to construct the variogram and to condition the geostatistical estimation.

Table 4 shows summary statistics for the upscaling errors. For all three transect designs, the mean error is slightly greater than zero indicating that the upscaling process slightly overestimates the true number of anomalies within each cell intersected by a transect. The median error across these same cells is exactly zero for all three transect designs.

Table 4. Summary statistics for the upscaling errors at all three transect spacings.

Statistic	105m Spacing	316m Spacing	631m Spacing
Mean	0.08	0.43	0.17
Std Dev	2.47	2.61	1.67
Min	-15.10	-14.57	-5.29
25th	-1.00	-0.14	-1.00
Median	0.00	0.00	0.00
75th	0.95	1.47	0.95
Max	14.24	11.67	8.91
Number	1083	343	170

The estimation errors are summarized in Table 5. The mean error is closest to zero for the 105m spacing transect design and is greater than zero for the 316m design and less than zero for the 631m spacing design. These results are consistent with the statistics on the sample data (Table 3) that showed the 105m spacing giving the results closest to the true data set and the 316 and 631m spacing data over and under estimating the true data values, respectively. The 25^{th} and 75^{th} percentiles of the error values in Table 5 show that for all transect designs, the number of anomalies within 50 percent of the cells is estimated to within +/- one anomaly. The largest errors in the estimated number of anomalies within a cell are +/- 15 to anomalies. The locations of these errors are examined below.

Table 5. Summary statistics for the estimation errors at the $1000m^2$ cell scale for all three transect spacings.

Statistic	105m Spacing	316m Spacing	631m Spacing
Mean	0.12	0.38	-0.21
Std Dev	2.43	2.89	2.81
Min	-18.67	-21.09	-24.10
25th	-0.76	-0.61	-0.95
Median	0.18	0.51	0.26
75th	1.17	1.38	1.13
Max	11.51	17.39	16.35
Number	4617	5357	5530

The scatterplots between the true and estimated number of anomalies within each cell are shown in Figure 8. These scatterplots provide a visual representation of the results summarized in Table 4 and 5. The scatterplots on the left side of Figure 8, with the red symbols, show the comparison for the upscaling process and the scatterplots on the right, blue symbols, show the comparison for the estimations. The degree to which the scatter of symbols follows the 1:1 line in each figure shows the amount of bias in the estimations. The upscaling process produces a relatively small amount of bias with the red symbols evenly surrounding the 1:1 line for all three transect designs. The estimation produces some amount of conditional bias, meaning that the amount of bias is dependent on the true value of the number of anomalies at each cell. To some extent the scatterplots show a tendency for the estimation process to overestimate low values and underestimate high values (blue symbols above the 1:1 line when the true value of the anomalies is low and below the 1:1 line when the true number of anomalies is high). This conditional bias is most obvious in the results of the 631m transect spacing (bottom, right scatterplot in Figure 8).

The total number of anomalies estimated for each transect design is calculated by summing the expected number of anomalies estimated at each cell. The totals of these sums are compared to the total true number of anomalies in Table 6. Results of this comparison show that the accuracy of the total anomaly estimate varies widely between the different transect designs. These results are consistent with the sample data obtained on the different transect designs as shown above in Table 3. The sample data for the 105m spacing is the most consistent with the true distribution of anomalies and, not surprisingly, the estimates made from these data provide the most accurate estimate of the total number of anomalies. The sample data for the 316m transect design overestimates the true distribution of anomalies and the sample data for the 631m transects underestimates the actual total number of anomalies.

Table 6. Results of the total estimated number of anomalies across the site for the three transect designs and compared to the true number of anomalies.

Transect Design	**105m**	**316m**	**631m**	**True**
Total Anomalies	16,679	18,241	14,878	16,033
Percent Error	4.0	13.8	-7.2	NA

The spatial distribution of the errors is shown in Figure 9. For all three transect designs, the majority of the site domain has errors near zero anomalies per cell. The major exception to this observation is the area of high anomaly intensity near the south central portion of the site. In this area, the anomaly estimation tends to over and under estimate the true number of anomalies per cell. The nature and degree of the over and underestimation changes between the different transect sampling designs. However, for all three transect designs, the largest degree of both over and under estimation occurs within this higher density region.

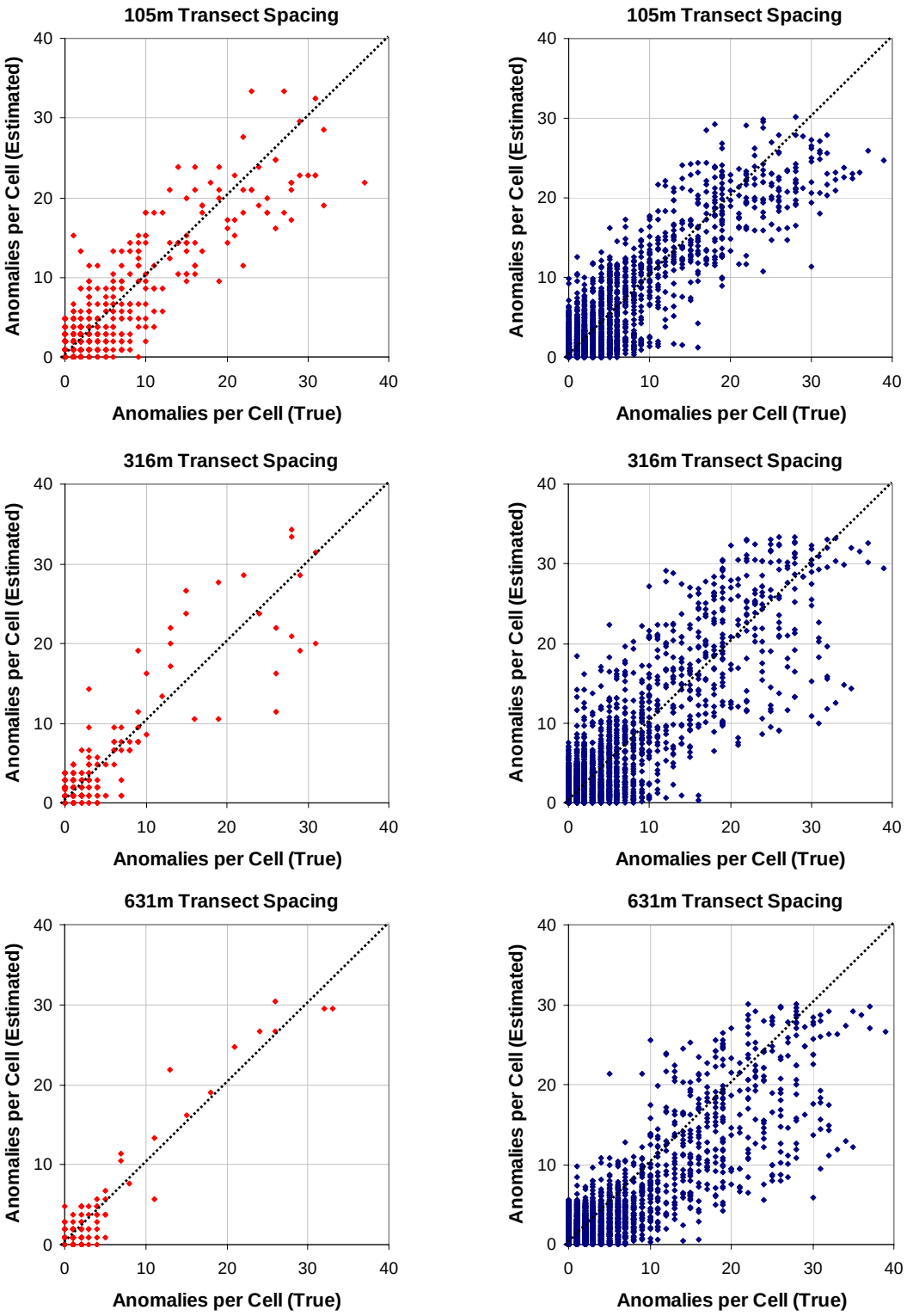

Figure 8. Comparison of the true and estimated number of anomalies per cell for the upscaled cells (left column) and the estimated cells (right column).

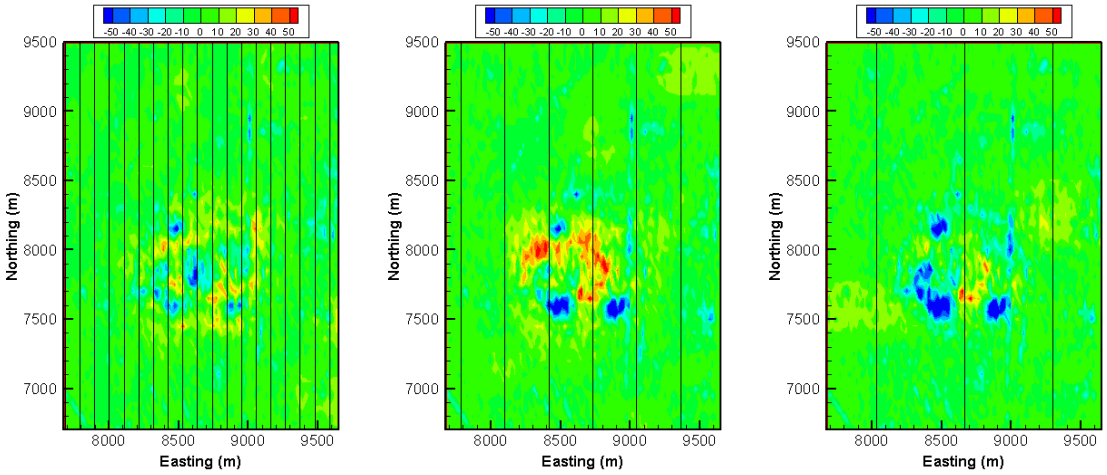

Figure 9. Spatial distribution of the combined errors (upscaling and estimation) for the three different transect designs. The color scale shows error in anomalies per acre for each 1000m^2 cell.

The maps in Figure 9 follow the results of the total anomaly estimation directly. The 105m spacing design produces the best estimate of the total number of anomalies and the map on the left of Figure 9 shows that the amounts of the largest over and under estimations (red and blue locations) are nearly equal. The locations of the largest over and under estimates are well mixed within the high anomaly density area. The 316m transect design overestimates the total number of anomalies (Table 3). The center map in Figure 9 shows that the number of overestimates within the high anomaly density area (red cells) dominates the number of underestimates (blue cells). The opposite is true for the 631m transect spacing as shown in the right map of Figure 9 and this is consistent with the overall underestimation of the number of anomalies by the 631m spacing transect design (Table 3).

The linear features of high anomaly intensity in the true data set are not well reproduced by the estimation procedure. For all three transect designs, these high intensity linear features are underestimated (linear arrangements of blue cells in Figure 9). This underestimation is not surprising given that the transects were not designed to detect these types of features and that these features are narrow and generally oriented parallel to the transects (north-south). For these features to be estimated accurately, it would be necessary for a transect to lie directly on top of them.

Summary

In summary, a five step process was developed to estimate the number of geophysical anomalies at each location across a site from limited transect sampling. The site area is discretized into a set of equal area cells, each 1000m^2, and anomaly estimates are made for each of these cells The anomaly estimation process includes an upscaling procedure to determine the representative number of anomalies in equal area cells crossed by the transects when the transects are considerably more narrow than the cells. These upscaled data are then used as input to a geostatistical estimation procedure that estimates the non-homogeneous intensity of the Poisson distribution defining the number of anomalies at all cells across the site. At this time, the number of anomalies within each cell is taken as the average of this distribution. Additional steps in the procedure necessary to provide a full distribution that defines the uncertainty in the estimated number of anomalies at each location are outlined.

This anomaly estimation procedure was applied to the Pueblo of Isleta S1 data set. Data were extracted from three different transect designs and then used as input to the anomaly estimation procedure. Results show that for all cases, the number of anomalies estimated at 50 percent of the locations was within +/- 1 anomaly of the true anomaly count. The maximum errors were +/- 15 to 20 anomalies. Keep in mind that these errors are per each 1000m^2 cell. The total number of anomalies estimated for the entire site is calculated as the sum of the estimated values across all cells. The total estimated number of anomalies was within +/- 15 percent of the true value for all three transect designs with the closest estimate of the total number of anomalies, off by 4 percent, resulting from the densest sample (105m spacing).

Acknowledgements

The authors thank Herb Nelson and Dan Steinhurst for providing the magnetometer data. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's national Nuclear Security Administration, under contract DE-AC04-94-AL-85000. This research is funded by the ESTCP program as part of project MM-0325.

Reference

Naval Research Laboratory, 2005, MTADS Airborne and Vehicular Survey of Target S1 at Isleta Pueblo, ESTCP Report, 45 pp.